Comparing the information conveyed by envelope modulation for speech intelligibility, speech quality, and music quality

James M. Kates[a] and Kathryn H. Arehart

Department of Speech Language and Hearing Sciences, University of Colorado, Boulder, Colorado 80309, USA

(Received 11 November 2014; revised 3 September 2015; accepted 9 September 2015; published online 27 October 2015)

This paper uses mutual information to quantify the relationship between envelope modulation fidelity and perceptual responses. Data from several previous experiments that measured speech intelligibility, speech quality, and music quality are evaluated for normal-hearing and hearing-impaired listeners. A model of the auditory periphery is used to generate envelope signals, and envelope modulation fidelity is calculated using the normalized cross-covariance of the degraded signal envelope with that of a reference signal. Two procedures are used to describe the envelope modulation: (1) modulation within each auditory frequency band and (2) spectro-temporal processing that analyzes the modulation of spectral ripple components fit to successive short-time spectra. The results indicate that low modulation rates provide the highest information for intelligibility, while high modulation rates provide the highest information for speech and music quality. The low-to-mid auditory frequencies are most important for intelligibility, while mid frequencies are most important for speech quality and high frequencies are most important for music quality. Differences between the spectral ripple components used for the spectro-temporal analysis were not significant in five of the six experimental conditions evaluated. The results indicate that different modulation-rate and auditory-frequency weights may be appropriate for indices designed to predict different types of perceptual relationships. © *2015 Acoustical Society of America.*

[http://dx.doi.org/10.1121/1.4931899]

[ICB] Pages: 2470–2482

I. INTRODUCTION

Envelope modulation is related to speech intelligibility (Drullman *et al.*, 1994; Xu and Pfingst, 2008), speech quality (van Buuren *et al.*, 1999), and music quality (Croghan *et al.*, 2014). The strength of these relationships has led to the development of several predictive indices based on measuring changes in the signal envelope modulation. However, the relative importance of different auditory frequency analysis bands and different modulation rate regions for modeling auditory judgments has not been established. This paper presents an analysis of the information provided by envelope modulation fidelity as a function of auditory analysis frequency and modulation rate when applied to speech intelligibility scores, speech quality judgments, and music quality judgments made by normal-hearing and hearing-impaired listeners.

The relationship between envelope modulation and speech intelligibility has been exploited in several intelligibility indices. The Speech Transmission Index (STI) (Steeneken and Houtgast, 1980; Houtgast and Steeneken, 1985), for example, uses bands of amplitude-modulated noise as probe signals and measures the reduction in signal modulation depth. Speech-based versions of the STI have been developed that are based on estimating the signal-to-noise ratio (SNR) from cross-correlations of the signal envelopes in each frequency band (Goldsworthy and Greenberg, 2004; Payton and Shrestha, 2013). Dubbelboer and Houtgast (2008) proposed using the SNR estimated in the modulation domain to estimate intelligibility, and Jørgensen and Dau (2011) and Chabot-Leclerc *et al.* (2014) have extended this concept to using the envelopes produced by a model of the auditory periphery (Dau *et al.*, 1997). In addition, Falk *et al.* (2010) have proposed using the ratio of low-to-high envelope modulation rate energies as a non-invasive intelligibility and quality index for reverberant speech. Taal *et al.* (2011) have developed the short-time objective intelligibility measure (STOI), which uses envelope correlations computed within auditory frequency bands for 382-ms speech segments, and an intelligibility index based on averaging envelope correlations for 20-ms speech segments has been developed by Christiansen *et al.* (2010). Changes in the signal spectro-temporal modulation, extracted from the envelope in each frequency band by forming a sequence of short-time spectra and then measuring how the spectral ripple fluctuates over time, have also been used to predict speech intelligibility (Chi *et al.*, 1999; Elhilali *et al.*, 2003; Kates and Arehart, 2014b).

Changes to the envelope have been used to predict speech and music quality as well. The HASQI speech quality index (Kates and Arehart, 2010, 2014a) starts with the short-time spectra produced at the output of an auditory model. At each segment interval, the spectra are fitted with half-cosine basis functions to produce a set of spectral ripple components. The quality index is computed using the normalized cross-covariances of the spectral ripple components of the

[a]Electronic mail: James.Kates@colorado.edu

degraded speech with those of the reference speech. In the PEMO-Q index (Huber and Kollmeier, 2006) that is used to predict both speech and music quality, the signal envelope is extracted from each frequency band in an auditory model. The envelope in each auditory frequency band is passed through a modulation filter bank. The degraded signal is compared to the clean reference signal by computing the normalized cross-correlation of each modulation rate output at each auditory analysis frequency and taking the average.

In some indices (e.g., Kates and Arehart, 2010, 2014a,b) the envelope modulation outputs are passed through a lowpass filter based on the temporal modulation transfer function (Viemeister, 1979; Dau *et al.*, 1997). The lowpass filter produces an implicit weighting of modulation rate by the envelope intensity within each modulation frequency region. In other indices, the outputs of the modulation filter bank, or the outputs produced by the spectro-temporal analysis, are combined across modulation filters. Typically, the normalized outputs of the modulation filters are summed across modulation rate using a uniform weighting (Steeneken and Houtgast, 1980; Elhilali *et al.*, 2003; Huber and Kollmeier, 2006; Jørgensen and Dau, 2011) independent of whether intelligibility or quality is being predicted. In the index developed by Falk *et al.* (2010), the four modulation rate bands below 22 Hz are combined with uniform weights to estimate the speech intelligibility or quality component of the degraded signal, while the four higher modulation rate bands are combined with uniform weights to give the degradation component. However, a recent paper by Chabot-Leclerc *et al.* (2014) has shown that for some signal degradations, improved performance in predicting intelligibility can be achieved by having the modulation filter weights depend on the signal characteristics. At this time there is no clear rationale for selecting one set of weights over another, so the first question considered in this paper is to determine which envelope modulation rates provide the most information relating signal characteristics to subject performance.

An additional consideration is the relative importance of the different auditory frequency bands or the different spectral ripple components. The STI (Steeneken and Houtgast, 1980; Houtgast and Steeneken, 1985) applies a set of auditory frequency-dependent weights to the modulation depth values measured within the separate auditory frequency bands. In most of the other indices based on the envelope outputs in auditory frequency bands (Huber and Kollmeier, 2006; Falk *et al.*, 2010; Christiansen *et al.*, 2010; Jørgensen and Dau, 2011; Taal *et al.*, 2011; Chabot-Leclerc *et al.*, 2014), a uniform weighting is applied to all of the auditory frequencies. Similarly, the intelligibility and quality indices developed by Kates and Arehart (2010, 2014a,b) apply uniform weights to the spectral ripple components. Thus a second question is to determine which auditory frequency bands provide the most information for the different types of stimuli.

A further consideration is the impact of hearing loss. Even when amplification is provided, hearing-impaired listeners often have more difficulty understanding speech presented in noise or modified by distortion than do normal-hearing listeners (Festen and Plomp, 1990). Hearing-impaired listeners also have difficulty in extracting the temporal fine structure of a signal (Hopkins *et al.*, 2008). And while the ability to use envelope modulation in understanding speech remains close to normal in quiet (Turner *et al.*, 1995; Lorenzi *et al.*, 2006), it is reduced for speech in noise (Başkent, 2006). However, quality ratings for speech and music after nonlinear and/or linear processing are similar for listeners with and without hearing loss (Arehart *et al.*, 2010, 2011). Thus a third objective of this paper is to compare the amount of information provided by changes in the envelope modulation in relation to perceptual judgements from normal-hearing (NH) and hearing-impaired (HI) listeners for the different types of stimuli.

The criterion used in this paper to evaluate the relative importance of the envelope modulation rate bands is mutual information (Moddemeijer, 1999; Taghia and Martin, 2014) and the uncertainty coefficient (Press *et al.*, 2007; Kates *et al.*, 2013). The uncertainty coefficient is used instead of the Pearson correlation coefficient. The models used to relate changes in the envelope modulation to the subject responses often involve nonlinear transformations of the measured envelope behavior; for example, the HASPI intelligibility index (Kates and Arehart, 2014b) uses a logistic transformation of the envelope modulation and the HASQI version 2 quality index (Kates and Arehart, 2014a) uses the square of the envelope modulation. The uncertainty coefficient gives the degree to which one variable is related to another without making any assumptions as to whether the relationship is linear or involves a nonlinear transformation, and it does not require that the form of the nonlinear transformation be known. The Pearson correlation coefficient, on the other hand, assumes a linear relationship between the two variables; transformed variables can be used in the Pearson correlation calculation, but the nature of the transformation must be specified prior to performing the calculation. Mutual information thus has the advantage of showing a potentially nonlinear relationship between variables without needing a mathematical expression of what that relationship may be, and mutual information can also describe nonlinear dependencies between variables that correlation may miss.

The purpose of this paper is to investigate the relative information provided by changes in envelope modulation for the following.

(1) Different procedures for describing the envelope modulation, specifically measuring the modulation rates within each auditory analysis band as compared to analyzing the spectro-temporal modulation across auditory bands.
(2) Different types of judgments, specifically quality as compared to intelligibility.
(3) Different stimuli, specifically quality ratings of music as compared to speech.
(4) Different hearing loss status, specifically listeners with sensorineural hearing loss as compared to listeners with normal hearing.

The remainder of the paper continues with a description of the subject data used for the intelligibility, speech quality, and music quality comparisons. The auditory model used to produce the envelope signals is presented next, followed by the procedures used to extract the envelope modulation within each auditory analysis band and to extract the spectro-temporal modulation across the auditory bands. The mutual information analysis and the uncertainty coefficient are then described. Results are presented showing the uncertainty coefficients between the envelope modulation rates and the subject intelligibility scores and quality ratings. The paper concludes with a discussion of the implications of the results for the design of intelligibility and quality indices.

II. METHODS

A. Subject data

The mutual information analysis is applied to datasets from several previously-published experiments. The datasets include speech intelligibility scores, speech quality ratings, and music quality ratings. The different datasets, summarized here, allow the comparison of the importance of auditory frequency bands and modulation rates for different types of experiments.

1. Speech intelligibility

Four intelligibility datasets were used in this paper. The same data were used by Kates and Arehart (2014b) in developing the HASPI intelligibility index, and an overview can be found in that paper. The data came from four experiments: (1) noise and nonlinear distortion (Kates and Arehart, 2005), (2) frequency compression (Souza *et al.*, 2013; Arehart *et al.*, 2013a), (3) noise suppression (Arehart *et al.*, 2013b), and (4) noise vocoder (Anderson, 2010).

The noise and nonlinear distortion experiment (Kates and Arehart, 2005) had 13 adult listeners with normal hearing and nine with mild-to-moderate sensorineural hearing loss. The test materials consisted of the Hearing-in-Noise-Test (HINT) sentences (Nilsson *et al.*, 1994). Each test sentence was combined with additive stationary noise, or was subjected to symmetric peak-clipping distortion or symmetric center-clipping distortion. The additive noise was extracted from the opposite channel of the HINT test compact disk. The peak-clipping and center-clipping distortion thresholds were set as a percentage of the cumulative histogram of the magnitudes of the signal samples for each sentence. Peak-clipping thresholds ranged from infinite clipping to no clipping, and center-clipping thresholds ranged from 98% to no clipping. The stimuli were presented to the normal-hearing listeners at a RMS level of 65 dB sound pressure level (SPL). The speech signals were amplified for the individual hearing loss, when present, using the National Acoustics Laboratories-Revised (NAL-R) linear prescriptive formula (Byrne and Dillon, 1986).

The frequency compression experiment (Souza *et al.*, 2013; Arehart *et al.*, 2013a) had 14 adult listeners with normal hearing and 26 listeners with mild-to-moderate sensorineural high-frequency loss. The stimuli for the intelligibility tests consisted of low-context IEEE sentences (Rosenthal, 1969) spoken by a female talker. The sentences were used in quiet and combined with multi-talker babble at SNRs ranging from −10 to 10 dB in steps of 5 dB. After the addition of the babble, the sentences were processed using frequency compression. Frequency compression was implemented using sinusoidal modeling (McAulay and Quatieri, 1986) in a two-channel implementation. The low frequencies passed through the system unaltered, while frequency compression was applied to the high frequencies. The ten highest peaks in the high-frequency band were selected, and the amplitude and phase of each peak were preserved while the frequencies were reassigned to lower values. Output sinusoids were then synthesized at the shifted frequencies (Quatieri and McAulay, 1986; Aguilera Muñoz *et al.*, 1999) and combined with the low-frequency signal. The frequency-compression parameters included three frequency compression ratios (1.5:1, 2:1, and 3:1) and three frequency compression cutoff frequencies (1, 1.5, and 2 kHz). The stimulus level for the normal-hearing subjects was 65 dB SPL. The speech signals were amplified for the individual hearing loss, when present, using NAL-R equalization (Byrne and Dillon, 1986).

The noise suppression experiment (Arehart *et al.*, 2013b) had seven younger adult listeners with normal hearing and thirty older adult subjects with mild-to-moderate sensorineural hearing loss. The stimuli consisted of low-context IEEE sentences (Rosenthal, 1969) spoken by a female talker. The sentences were combined with multi-talker babble at signal-to-noise ratios of −18 to +12 dB in steps of 6 dB. The sentence level prior to noise suppression was set to 65 dB SPL. The noisy speech stimuli were processed with an ideal binary mask noise-reduction strategy (Kjems *et al.*, 2009). The processing was implemented using 20-ms time frames having 50% overlap. The local SNR was computed for each time-frequency cell and compared to a local criterion (LC) of 0 dB, resulting in a gain decision of 1 if the local SNR was above LC, and 0 otherwise. Similar to the procedure in Li and Loizou (2008), errors were introduced into the ideal binary mask by randomly flipping a certain percentage (0, 10, and 30%) of the gain decisions for each time-frequency cell either from 0 to 1 or from 1 to 0. The binary patterns were then converted into gain values, where 1's were converted into 0 dB gain and the zeros were converted into an attenuation of either 10 or 100 dB. Following the noise suppression processing, the speech signals were amplified for the individual hearing loss, when present, using NAL-R equalization (Byrne and Dillon, 1986).

The noise vocoder experiment (Anderson, 2010) had ten adult subjects with normal hearing and ten with mild-to-moderate sensorineural hearing loss. The test materials were low-context sentences from the IEEE corpus (Rosenthal, 1969) spoken by a male and by a female talker. The speech was processed without any interfering signal or combined with multi-talker babble at SNRs of 18 and 12 dB. The sentences were passed through a bank of 32 band-pass filters with center frequencies distributed on an auditory frequency scale. The speech envelope in each band was extracted via the Hilbert transform followed by a 300-Hz lowpass filter. Two types of vocoded signals using noise carriers were produced. One signal was produced by multiplying the noise carrier in each frequency band by the speech envelope determined for that band (Shannon *et al.*, 1995). For the second signal, the fluctuations of the noise carrier within the frequency band were first reduced by dividing the noise carrier by its own envelope (Kohlrausch *et al.*, 1997) before multiplying it by the speech envelope for the band. The vocoding was applied to the speech-plus-babble signal starting with the highest frequency bands and proceeding to lower frequencies. The degree of vocoding was increased in steps of two frequency bands from no bands vocoded to the 16 highest-frequency bands vocoded. The stimulus level for the normal-hearing listeners was 65 dB SPL, and NAL-R amplification (Byrne and Dillon, 1986) was provided for the HI listeners.

2. Speech quality

The speech quality data used in this paper comprised the noise and nonlinear distortion results reported by Arehart *et al.* (2010). A total of 14 subjects with normal hearing and 15 subjects with mild to moderate-severe sensorineural hearing losses took part in the experiment. The test materials were two sets of concatenated sentences from the HINT (Nilsson *et al.*, 1994), with one concatenated two-sentence set spoken by a male talker and another two-sentence set spoken by a female talker (Nilsson *et al.*, 2005). The task of the listener was to indicate the rating of sound quality on a rating scale which ranged from 1 (poor sound quality) to 5 (excellent sound quality) (International Telecommunication Union, 2003).

In previous analyses (Kates and Arehart, 2010, 2014a) using these data, the quality ratings for each subject were normalized so that the highest observed rating was reset to 1 and the lowest rating was reset to 0. The rating normalization reduced the intersubject variability caused by different subjects adopting different internal anchors or using only part of the rating scale. In the present study, the differences in the mutual information analysis for the normalized versus unnormalized data were very small, and the results for the unnormalized data are presented.

The processing conditions were implemented using a simulated hearing aid programmed in MATLAB. The order of processing was additive noise, followed by nonlinear processing. The final processing step was to adjust the loudness of the filtered signal to match that of the unprocessed reference. The level of presentation for the subjects in the NH group was 72 dB SPL, and the stimuli were amplified for listeners in the HI group using the NAL-R linear prescriptive formula based on individual thresholds (Byrne and Dillon, 1986). Stimuli were presented to the listeners monaurally using headphones in a sound booth.

The noise conditions included speech in stationary speech-shaped noise and speech in multi-talker babble at a range of SNRs. The distortion conditions included instantaneous peak clipping and amplitude quantization. Dynamic-range compression (Kates and Arehart, 2005) was included both for the clean speech and for speech in babble. Spectral subtraction (Tsoukalis *et al.*, 1997) was included to give a set of noise-suppression conditions for speech in babble at a range of SNRs. The final noise and nonlinear condition combined spectral subtraction with compression for speech in babble.

3. Music quality

The music quality data used in this paper comprised the noise and nonlinear distortion results reported by Arehart *et al.* (2011). A total of 19 subjects with normal hearing and 15 subjects with mild to moderate-severe sensorineural hearing losses took part in the experiment. Three music segments, each of approximately 7 s duration, were used. The first segment was an excerpt from a jazz trio comprising piano, string bass, and drums. The second segment was an excerpt from the second movement of Haydn's Symphony No. 82, which features a full orchestra. The third segment was an extract of a jazz vocalist singing nonsense syllables ("scat" singing) without any accompaniment. The stimulus presentation and signal processing conditions of the music experiment duplicated those of the speech quality experiment described above, and unnormalized quality ratings are used for the analysis presented in this paper.

B. Auditory model

The envelope signals analyzed in this paper were the outputs from an auditory model. A detailed description of the model is presented in Kates (2013), and summaries are presented in Kates and Arehart (2014a,b). The envelope modulation analysis compares the model outputs for two separate signals: one set of outputs is for the reference signal that is free of any degradation, while the second set of outputs is for the degraded signal. The signal comparison uses normalized cross-covariances of the envelope signals. For NH listener calculations, both the reference and degraded signal outputs are produced using the auditory model adjusted for normal hearing. For HI listener intelligibility calculations, the reference signal is passed through the model adjusted for the normal auditory periphery while the degraded signal is passed through the model adjusted for the impaired ear. For HI quality calculations, both signals are passed through the model adjusted to reproduce the impaired periphery.

The auditory model is shown in the block diagram of Fig. 1. The model can be adjusted to reflect the effects of outer hair-cell (OHC) and inner hair-cell (IHC) damage, and the model for normal hearing is the same as for hearing loss but with the OHC and IHC damage set to zero. The model is designed to reproduce results from headphone listening, so the head-related transfer function and ear-canal resonance are not included. The model operates at a 24-kHz sampling rate, and the signal is first resampled if needed. The resampling is followed by the middle ear filter, which is implemented as a two-pole highpass filter at 350 Hz in series with a one-pole lowpass filter at 5000 Hz. The auditory analysis uses a 32-band gammatone filter bank with band center frequencies spanning 80 to 8000 Hz, and the filter bandwidths are increased in response to increasing signal intensity. The dynamic-range compression associated with the OHC function is implemented by multiplying the auditory filterbank

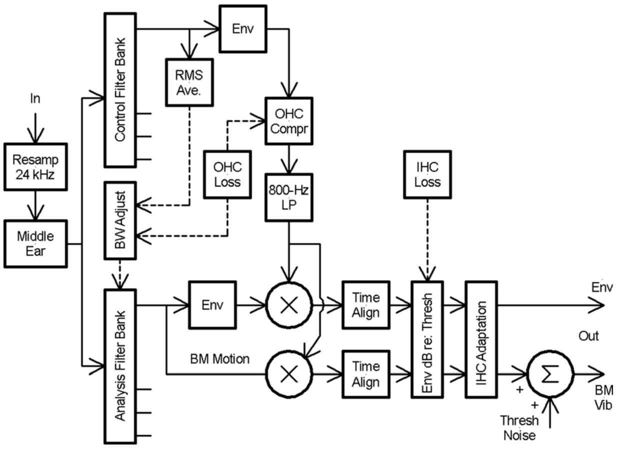

FIG. 1. Block diagram of the auditory model used for the envelope analysis.

output by a control signal produced by the control filter bank. The control signal provides linear amplification for inputs below 30 dB SPL and above 100 dB SPL, and provides frequency-dependent compression in between. The compression ratio ranges from 1.25:1 at 80 Hz to 3.5:1 at 8000 Hz for the normal ear. The control filters are set to the widest bandwidths allowed by the model, which also produces two-tone suppression. IHC firing rate adaptation, with time constants of 2 and 60 ms, is the final processing stage in the model. Hearing loss is represented by an increase in the auditory filter bandwidth, a reduction in the OHC compression ratio, a reduction in two-tone suppression, and a shift in auditory threshold. Two model outputs are available, one which includes the signal temporal fine structure and one which tracks the signal envelope; only the envelope output has been used in this paper.

C. Envelope modulation

Two envelope analysis procedures are used in this paper. One compares the envelope modulation within auditory frequency bands, while the second compares spectro-temporal modulation across auditory frequency bands as measured using cepstral correlation coefficients (Kates and Arehart, 2010, 2014a,b). For both procedures, the envelope modulations of the degraded signal are compared to those of the clean reference using a normalized cross-covariance. Perfect agreement yields a value of 1, while completely independent envelope modulation in the two signals produces a value of 0.

The auditory band envelope modulation comparison is shown in Fig. 2. The output of the auditory model was the envelope in each frequency band after being converted to dB re: auditory threshold. The envelopes in each frequency band were lowpass filtered at 320 Hz using a 384-tap (16 ms) linear-phase finite-impulse response (FIR) filter, and the envelope was resampled at 1000 Hz. The intervals in the stimuli falling below a silence threshold were pruned, and the envelopes then passed through a modulation filterbank comprising ten filters from 0 to 320 Hz implemented using 512-tap linear-phase FIR filters at the 1-kHz sampling rate. The leading and trailing filter transients were removed, giving filtered envelopes having the same length as the input envelope sequences. The ten modulation filter bands are listed in Table I. For each modulation filter output in each auditory analysis band, the envelope of the processed signal being evaluated was compared to the envelope of the unprocessed reference signal using a normalized cross-covariance, giving a value between 0 and 1, with 1 representing perfect envelope fidelity. The result of the envelope modulation analysis was a correlation matrix having 32 auditory analysis bands by ten envelope modulation rate filters.

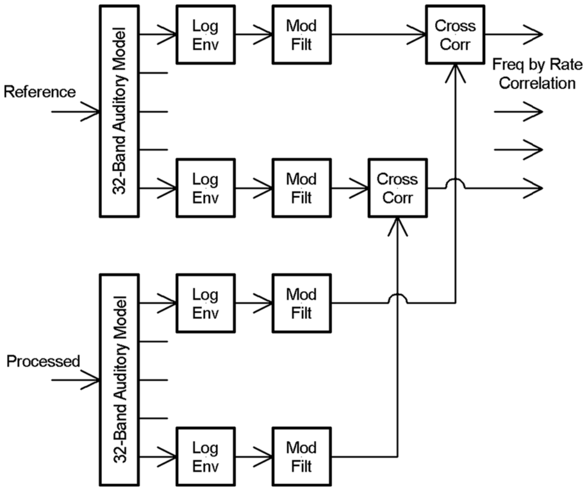

FIG. 2. Block diagram showing the envelope modulation cross-correlation procedure.

The spectro-temporal modulation comparison procedure is shown in Fig. 3. The processing started with the same filtered and sub-sampled log-envelope signals as used for the envelope comparison in auditory filter bands, again with the sub-threshold segments removed. At each time increment at the 1-kHz sub-sampling rate, the log-envelope samples in the 32 auditory analysis filters were fitted with a set of ten basis functions that started with one-half cycle of a cosine spanning the spectrum and extended to five cycles spanning the spectrum. The spectrum consists of log amplitude values in each auditory frequency band. The basis functions were thus fit to a log spectrum computed on an auditory frequency scale, so they correspond to short-time mel cepstral coefficients. More details on the calculation of the cepstral coefficients can be found in Kates and Arehart (2010, 2014a,b). The basis functions also correspond to the principal components of speech

TABLE I. Modulation rate filters used for the envelope analysis.

Filter number	Modulation filter range, Hz
1	0–4
2	4–8
3	8–12.5
4	12.5–20
5	20–32
6	32–50
7	50–80
8	80–125
9	125–200
10	200–325

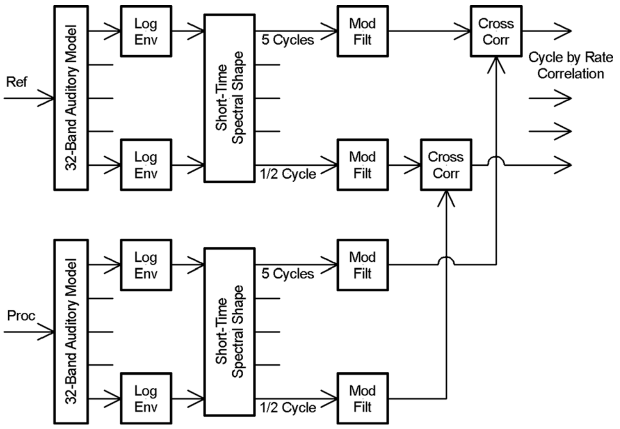

FIG. 3. Block diagram showing the cepstral correlation procedure.

determined by Zahorian and Rothenberg (1981), who found that the first five principal components explained about 90% of the speech short-time spectral variance and that the first ten components explained about 97% of the variance. The ten cepstral coefficients were then passed through the modulation filterbank. For each modulation filter output for each cepstral coefficient, the envelope of the processed signal being evaluated was compared to the envelope of the unprocessed reference signal using a normalized cross-covariance. The result of the envelope modulation analysis was a correlation matrix having ten cepstral correlation coefficients by ten envelope modulation rate filters.

Examples of the normalized cross-covariances are presented in Fig. 4. The speech stimulus was a pair of sentences spoken by a female talker, and the auditory model parameters were set for normal hearing. The interference in plots (a) and (c) was multi-talker babble at a signal-to-noise ratio (SNR) of 20 dB, and the SNR was reduced to 5 dB for plots (b) and (d). Plots (a) and (b) are for the auditory band correlation analysis illustrated in Fig. 2, while plots (c) and (d) are for the cepstral correlation analysis illustrated in Fig. 3. For the auditory band procedure, the normalized cross-correlation tends to be highest at low envelope modulation rates and high auditory frequencies. For the cepstral correlation procedure, the normalized cross-correlation tends to be highest at low modulation rates, with a relatively weak dependence on spectral ripple density. For both procedures, reducing the SNR reduces the magnitude of the normalized cross-correlation while preserving the general pattern across modulation rate and auditory frequency.

D. Uncertainty coefficients

The uncertainty coefficient (Press *et al.*, 2007; Kates *et al.*, 2013) is the ratio of the mutual information between

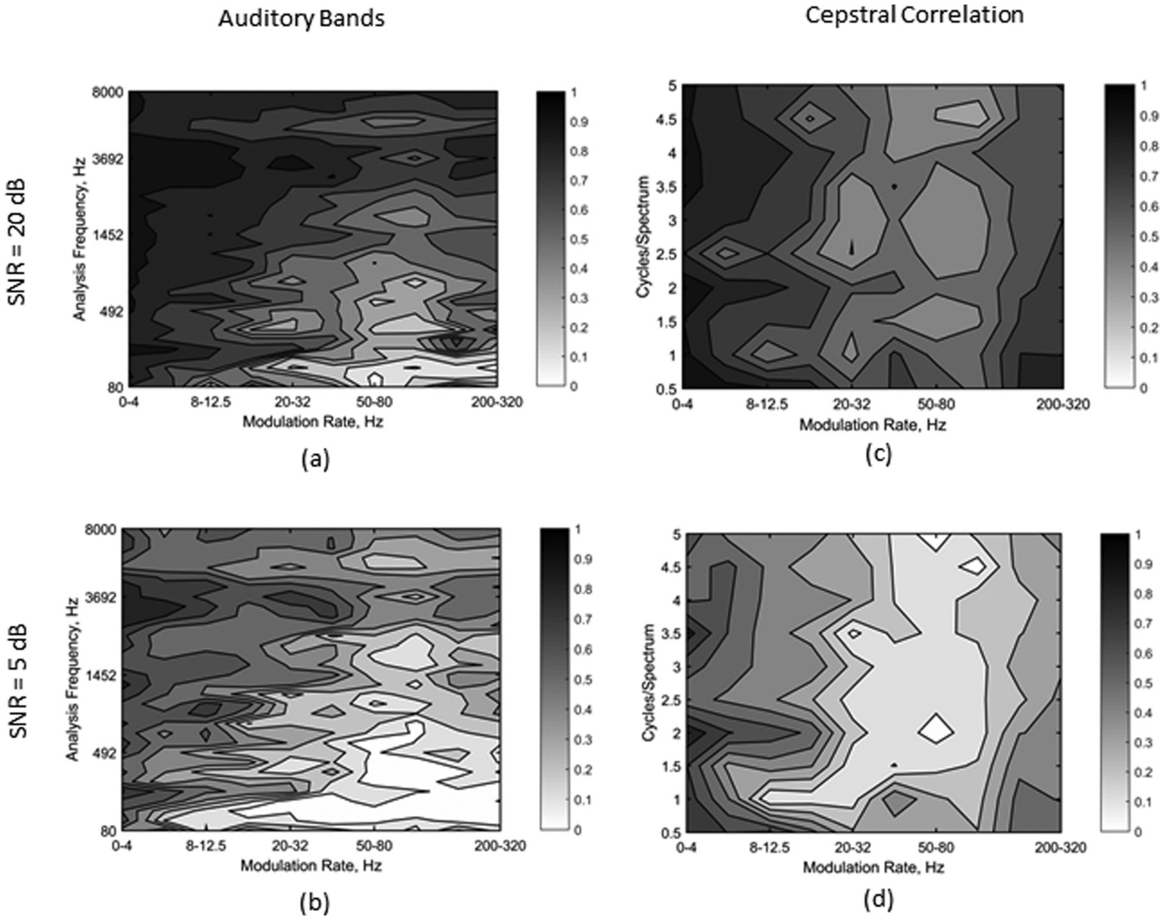

FIG. 4. Contour plots showing the normalized envelope cross-correlations for a pair of sentences spoken by a female talker in a background of multi-talker babble for normal hearing: (a) envelope modulation rate in each auditory analysis band, SNR = 20 dB, (b) envelope modulation rate, SNR = 5 dB, (c) cepstral correlation, SNR = 20 dB, and (d) cepstral correlation, SNR = 5 dB.

two variables to the entropy of one or both of those variables. The mutual information $I(x,y)$ between random variables x and y is given by the entropies $H(x)$, $H(y)$, and $H(x,y)$:

$$I(x;y) = H(x) + H(y) - H(x,y). \tag{1}$$

The entropy measures the uncertainty or randomness of a variable. The entropy in bits is given by

$$H(x) = -\sum_{x} P(x)\log_2 P(x), \tag{2}$$

where $P(x)$ is the probability density function for x and the summation is over all observed values of x. The uncertainty coefficient is then calculated as

$$U(x,y) = I(x;y)/H(x). \tag{3}$$

A more comprehensive explanation of mutual information and the uncertainty coefficient is provided by Kates *et al.* (2013).

The uncertainty coefficient was used to relate the modulation filter cross-covariance values to the subject results. The envelope modulation correlations and cepstral correlations between the degraded and reference signals were computed for each processing condition and subject in each of the experiments. Twelve separate sets of comparisons were computed based on the type of correlation (auditory band modulation correlation or cepstral correlation), type of experiment (intelligibility, speech quality, or music quality), and hearing loss group (normal or impaired hearing). For each of these twelve sets of conditions, the mutual information was computed between the subject results (intelligibility score or quality rating) and the normalized envelope cross-covariance value for each combination of auditory analysis frequency (or cepstral correlation coefficient) and envelope modulation rate. The entropy of the subject results was also computed for the data, and the uncertainty coefficient at each combination of auditory frequency and modulation rate was produced by dividing the mutual information by the subject entropy.

III. RESULTS

The uncertainty coefficients for datasets from the NH and HI listeners are plotted in Figs. 5 and 6, respectively. Each of the two figures comprises six sub-plots. Plots (a)–(c) are for the auditory band correlation analysis, while plots (c)–(e) are for the cepstral correlation analysis. Plots (a) and (d) are for speech intelligibility, plots (b) and (e) are for speech quality, and plots (c) and (f) are for music quality. Each sub-plot shows the information measured between the envelope modulation correlations and the subject data as a function of envelope modulation rate and auditory frequency or spectral ripple.

None of the uncertainty coefficient distributions satisfied the Kolmogorov-Smirnov goodness-of-fit hypothesis test, so the nonparametric Friedman test was used to determine if auditory frequency or envelope modulation rate were significant factors. The results of the Friedman test are presented in Table II. The modulation rate is significant for all twelve information analyses: both hearing groups, all three experiments and both envelope analysis procedures. When the envelope analysis is performed in auditory frequency bands, the frequency band is significant for both listener groups and all three experiments. However, the cepstral correlation basis function (spectral ripple density) is only significant for the HI speech quality, and is not significant for the other five comparisons.

The Tukey HSD test was used to identify significant differences in auditory frequency band and envelope modulation rate. In Fig. 5(a) for NH intelligibility, the uncertainty coefficients for modulation rates below 12.5 Hz are significantly greater than for the rates above 20 Hz, and the coefficients for auditory frequencies between 1.6 and 2.1 kHz are significantly greater than the coefficients for frequencies between 4.1 to 5.2 kHz and between 6.4 to 8 kHz. In Fig. 5(d), the uncertainty coefficients for modulation rates below 12.5 Hz are significantly greater than the coefficients between 32 and 80 Hz and between 125 and 320 Hz.

In Fig. 5(b) for NH speech quality, the uncertainty coefficients for modulation rates below 12.5 Hz are significantly lower than the coefficients for rates between 20 and 320 Hz, and the uncertainty coefficients for modulation rates between 12.5 and 20 Hz are significantly lower than the coefficients for rates between 50 and 80 Hz and between 125 and 320 Hz. The uncertainty coefficients for auditory frequency bands between 80 and 200 Hz are significantly lower than the coefficients between 1.5 and 2.6 kHz. In Fig. 5(e), the uncertainty coefficients for modulation rates less than 4 Hz are significantly lower than the coefficients for rates between 125 and 320 Hz, and the coefficients for rates below 4 Hz and between 8 and 20 Hz are significantly lower than the coefficient for rates between 125 and 200 Hz.

In Fig. 5(c) for NH music quality, the uncertainty coefficients for modulation rates below 12.5 Hz are significantly lower than the coefficients for rates between 20 and 320 Hz, and the coefficients for rates less than 20 Hz are significantly lower than the coefficients for rates between 20 to 80 Hz and between 125 and 320 Hz. The uncertainty coefficients for auditory frequency bands between 80 and 490 Hz are significantly lower than the coefficients for frequency bands between 2.6 and 7.2 kHz. In Fig. 5(f), the uncertainty coefficients for modulation rates less than 4 Hz are significantly lower than the coefficients for rates above 20 Hz.

The patterns for the HI listeners are similar to those for the NH listeners. In Fig. 6(a) for HI intelligibility, the uncertainty coefficients for modulation rates between 8 and 12.5 Hz and between 20 and 50 Hz are significantly greater than the coefficients for rates between 125 and 320 Hz. The uncertainty coefficients for auditory bands between 80 and 240 Hz are significantly greater than the coefficients for bands between 4.6 and 7.2 kHz, the coefficients for auditory bands between 870 Hz and 1.3 kHz are significantly greater than the coefficients for bands between 4.1 and 8 kHz, and the uncertainty coefficients for auditory bands between 420 Hz and 1.3 kHz are significantly greater than the coefficients for bands between 5.2 and 8 kHz. In Fig. 6(d), the

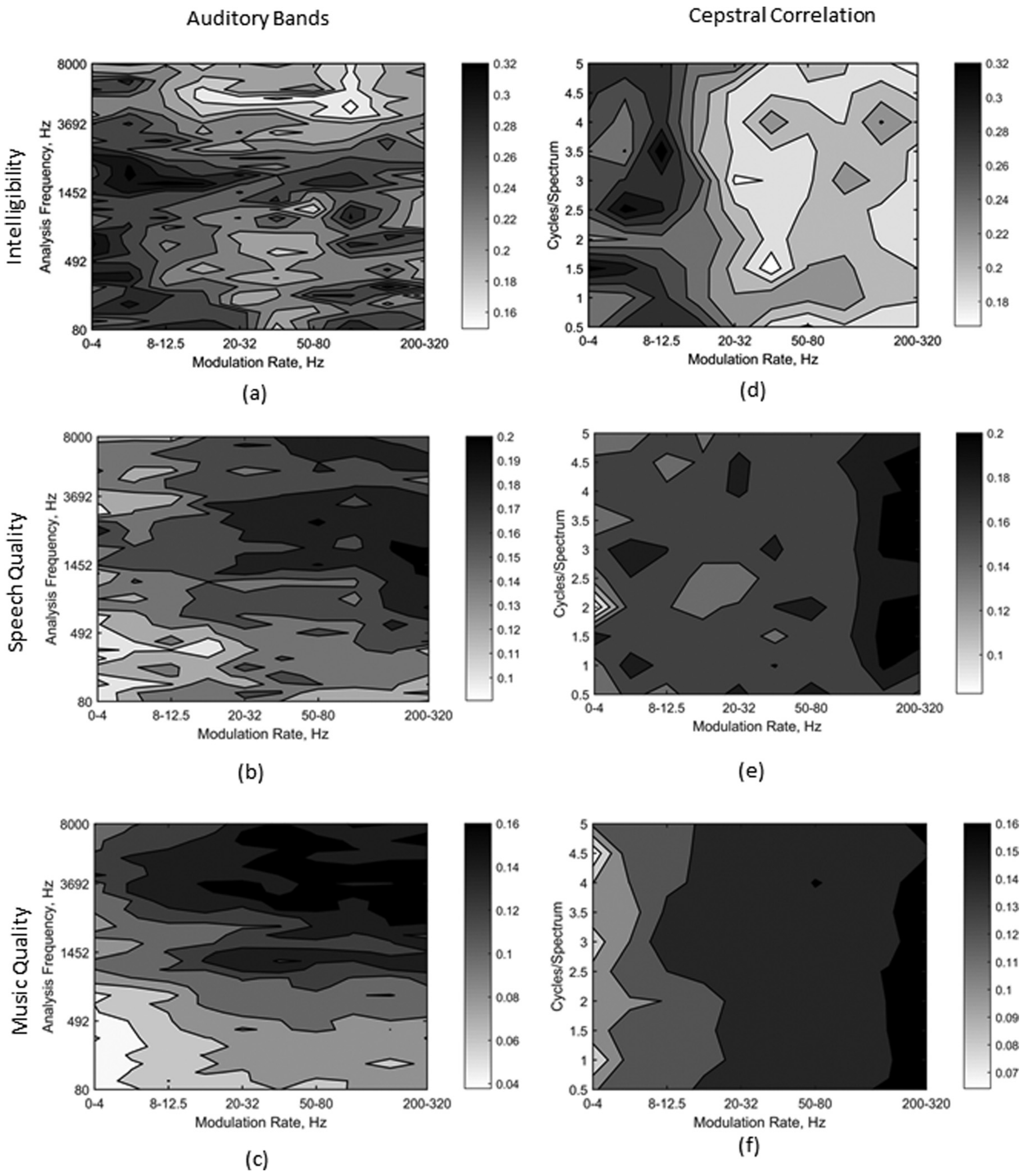

FIG. 5. Contour plots showing the uncertainty coefficients between the envelope information and the subject intelligibility scores or quality ratings for NH listeners. The left column (a)–(c) is for envelope modulation rate measured in each auditory frequency band, while the right column (d)–(f) is for the short-time spectrum fit with the cepstral correlation basis functions. The top row (a), (d) shows sentence intelligibility, the second row (b), (e) shows sentence quality, and the third row (c), (f) shows music quality.

uncertainty coefficients for modulation rates between 8 and 12.5 Hz and between 20 and 50 Hz are significantly greater than the coefficients for rates between 125 and 320 Hz.

In Fig. 6(b) for HI speech quality, the uncertainty coefficients for modulation rates below 8 Hz are significantly lower than the coefficients for rates above 20 Hz, and the uncertainty coefficients for modulation rates below 12.5 Hz are significantly lower than the coefficients for rates between 20 and 80 Hz and between 125 and 320 Hz. The uncertainty coefficients for auditory bands between 80 and 150 Hz are significantly lower than the coefficients for bands between 570 and 660 Hz, and the coefficients for bands between 5.7 and 8 kHz are significantly lower than for the bands between 490 and 760 Hz, between 1.3 and 1.6 kHz, and between 2.1 and 2.6 kHz

In Fig. 6(e), the uncertainty coefficients for modulation rates below 4 Hz are significantly lower than the coefficients for rates between 32 and 50 Hz and between 125 and 320 Hz, and coefficients for rates below 8 Hz are significantly lower than the coefficients for rates between 125 and 200 Hz. The uncertainty coefficients for cepstral correlation basis function 2 (1 cycle/spectrum) are significantly lower than the coefficients for basis functions 6 through 8 (3 to 4 cycles/spectrum).

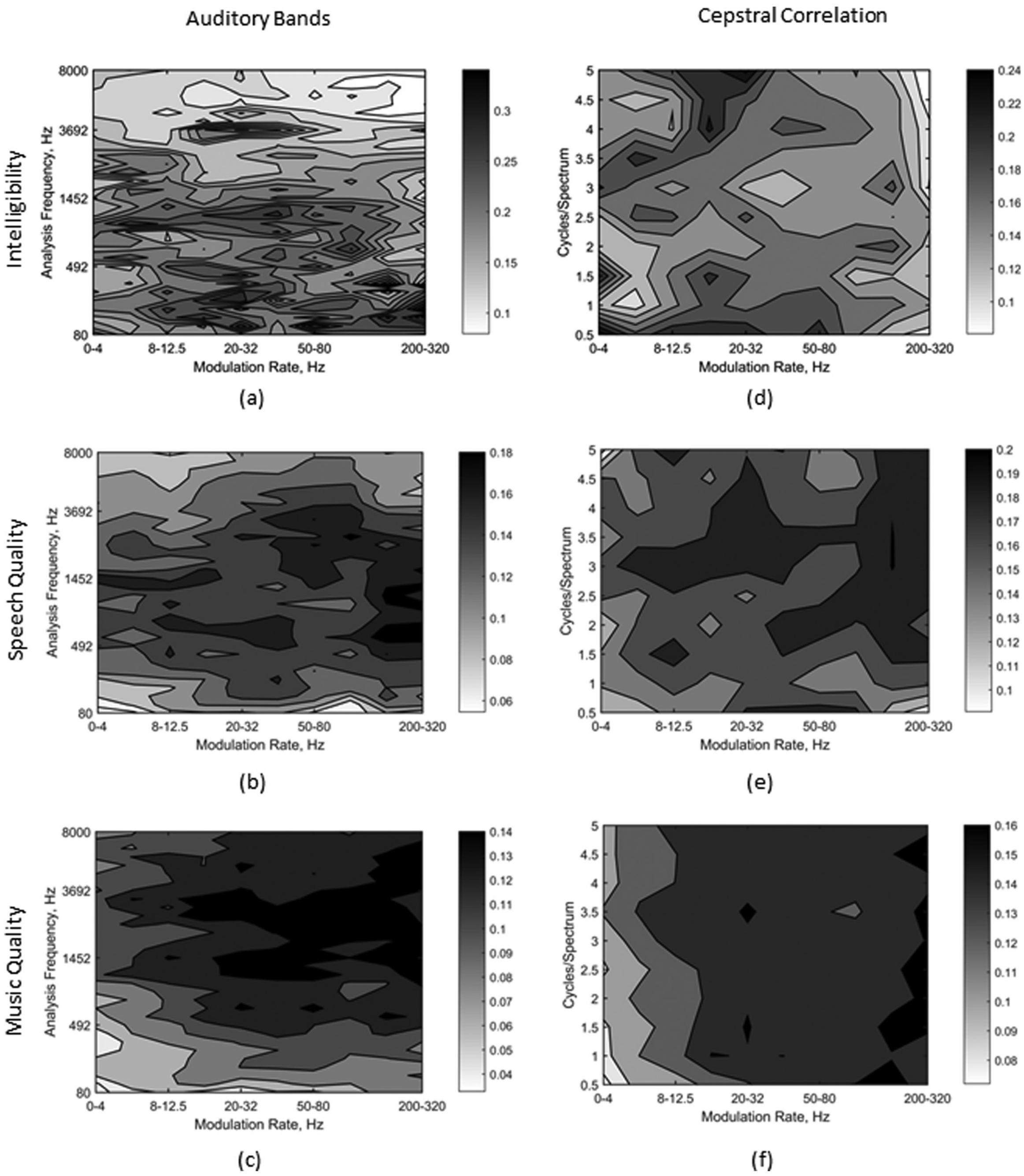

FIG. 6. Contour plots showing the uncertainty coefficients between the envelope information and the subject intelligibility scores or quality ratings for hearing-impaired (HI) listeners. The arrangement is the same as for Fig. 5.

In Fig. 6(c) for HI music quality, the uncertainty coefficients for modulation rates below 8 Hz are significantly lower than the coefficients for rates above 12.5 Hz, and the coefficients for rates below 12.5 Hz are significantly lower than the coefficients for rates above 20 Hz. The uncertainty coefficients for auditory bands between 80 and 350 Hz are significantly lower than the coefficients for bands between 1.1 and 3.7 kHz and at 4.6 kHz. The uncertainty coefficients for the auditory bands at 1.5 kHz is significantly greater than the coefficient for the band at 8 kHz. In Fig. 6(f), the uncertainty coefficients for modulation rates below 8 Hz are significantly lower than the coefficients for rates between 20 and 50 Hz and between 125 and 320 Hz, and the coefficients for rates below 12.5 Hz are significantly lower than the coefficients for rates between 125 and 320 Hz.

IV. DISCUSSION

The results show that envelope modulation rate is a significant factor for both the auditory band and cepstral correlation analysis procedures. The auditory frequency band is also significant, but the spectral ripple density is in general not significant. The patterns of the contour plots indicate that lower envelope modulation rates (below 12.5 Hz) are most important for intelligibility, while higher modulation rates (above 20 Hz) are most important for speech and music

TABLE II. Friedman test results for the significant factors in each of the uncertainty coefficient contour plots shown in Figs. 5 and 6. NH refers to normal-hearing and HI refers to hearing-impaired listeners.

Subject group	Signal property	Envelope analysis	Modulation rate		Auditory frequency	
			Chi-Sq.	Prob.	Chi-Sq.	Prob.
NH	Intelligibility	Aud. band	105.7	<0.001	140.4	<0.001
		Cep. corr.	68.03	<0.001	4.67	0.862
	Speech quality	Aud. band	153.3	<0.001	191.5	<0.001
		Cep. corr.	42.59	<0.001	14.42	0.108
	Music quality	Aud. band	211.0	<0.001	285.9	<0.001
		Cep. corr.	77.24	<0.001	8.29	0.505
HI	Intelligibility	Aud. band	38.76	<0.001	210.3	<0.001
		Cep. corr.	26.14	<0.001	9.08	0.430
	Speech quality	Aud. band	93.87	<0.001	207.8	<0.001
		Cep. corr.	34.69	<0.001	28.36	<0.001
	Music quality	Aud. band	190.7	<0.001	246.8	<0.001
		Cep. corr.	66.81	<0.001	5.06	0.829

quality. The patterns also indicate that low-to-mid auditory frequencies (between 80 Hz and 2 kHz) are most important for intelligibility, while mid frequencies (between 500 and 2500 Hz) are most important for speech quality and higher frequencies (between 2.5 and 8 kHz) are most important for music quality. Spectral ripple components between 3 and 4 cycles/spectrum are most important for speech quality judgments made by hearing-impaired listeners, but otherwise spectral ripple is not a significant factor.

The normalized envelope cross-covariance used in this paper measures envelope fidelity. This measurement is distinct from envelope intensity since it is possible to have a high degree of envelope similarity even in modulation rate regions where the modulation intensity is low. The distinction is especially important for music quality, where the high modulation rate bands conveyed the most information about the music quality ratings even though the intensity of the envelope modulations at high modulation frequencies is low (Croghan *et al.*, 2014).

The cross-correlation metric is also related to the modulation SNR estimated from the envelopes of the noisy speech and noise signals within auditory frequency bands (Dubbelboer and Houtgast, 2008; Chabot-Leclerc *et al.*, 2014). Benesty *et al.* (2008) have shown that for a signal corrupted by additive noise, the correlation coefficient between the clean signal and the noisy signal is related to the SNR

$$\rho^2 = SNR/(1+SNR). \tag{4}$$

Since the mutual information is unaffected by signal transformations, the information conveyed by the modulation SNR will be similar to the information conveyed by the modulation correlation coefficient. Thus the relative information provided by the different modulation filter bands measured in this paper would also apply to intelligibility and quality indices based on the modulation SNR.

The emphasis on low modulation rates suggested by the uncertainty coefficients for speech intelligibility differs from the weights proposed by Chabot-Leclerc *et al.* (2014), who for each sentence placed increased weight on the modulation filter bands having the greatest modulation SNR variance across auditory analysis frequency. Chabot-Leclerc *et al.* (2014) thus modified their weights for each stimulus and type of signal degradation they considered. Their weighting scheme improved the accuracy of the intelligibility predictions for speech corrupted by timing jitter, but reduced the accuracy for reverberant speech in comparison with using uniform weights at all modulation rates. The uncertainty coefficients in this paper represent the average computed over several experiments, which will reduce the potential dependence on the relative information associated with any particular signal processing system. The uncertainty coefficients thus show the general trends in the importance of the different modulation rates over a wide variety of processing conditions.

Envelope modulation rates above 125 Hz are generally not used in calculating speech intelligibility or quality indices due to the reduced auditory sensitivity to amplitude modulation at higher modulation rates (Viemeister, 1979; Dau *et al.*, 1997). Viemeister (1979), for example, found that amplitude modulation sensitivity for a broadband noise carrier could be modeled by a lowpass filter having a cutoff frequency of 65 Hz. However, in this study envelope modulation rates between 125 and 320 Hz were shown to be significant for speech quality ratings made by both NH and HI listeners. This range of modulation rates corresponds to the pitch periodicity of human speech (Schwartz and Purvis, 2004), so a possible explanation is that noise and nonlinear distortion may affect the periodicity of the speech by adding timing jitter or generating inharmonic spectral components. These periodicity changes could also be related to an increase in auditory roughness (Terhardt, 1974; Zwicker and Fastl, 1999; Tufts and Molis, 2007). Roughness is often reported as a harshness or raspiness of the speech. Roughness is maximum for modulation rates in the vicinity of 75 Hz, and is perceptible out to a modulation rate of about 300 Hz.

The majority of intelligibility and quality indices average over the envelope modulation rates using a uniform weighting, which also means that the same modulation weights are used for quality as are used for intelligibility.

The results of the analysis in this paper indicate that uniform weights do not accurately reflect the relative information provided by different modulation rates for intelligibility *versus* quality, or for modeling quality ratings for music as opposed to speech. Given the results of this paper, the uniform weighting commonly used appears to be a compromise between the low modulation rates that provide the most information for intelligibility and the high modulation rates that provide the most information for quality. Using the same modulation weights for both intelligibility and speech quality, or for both speech and music quality, would thus be expected to lead to sub-optimal solutions since the information provided by different modulation rates differs for the different problems.

The majority of intelligibility and quality indices also average over the auditory frequency bands or spectral ripple components using uniform weights. The results reported in this paper show that auditory frequency band is a significant factor, but that spectral ripple component generally is not. Thus, for an intelligibility or quality index based on envelope modulation within auditory frequency bands, there may be an advantage to determining separate weights for each combination of auditory analysis frequency and modulation rate. For indices based on cepstral correlation, on the other hand, one can average over the spectral ripple components without any apparent loss of information.

The maximum spectral ripple density of five cycles/spectrum used in this study is lower than the two cycles/octave auditory spectral resolution that appears to be necessary for accurate speech recognition (van Veen and Houtgast, 1985; Henry *et al.*, 2005). However, the objective of the cepstral correlation measurements in this paper is to measure changes in the signal time-frequency modulation that are related to changes in speech intelligibility or quality. For example, both broadband additive noise and multichannel dynamic-range compression reduce spectral contrast. The associated changes in the short-time spectra, as measured using cepstral correlation coefficients spanning 0.5 to 2.5 cycles/spectrum, are highly correlated with listener results for intelligibility (Kates and Arehart, 2014b) and speech quality (Kates and Arehart, 2010, 2014a). Furthermore, the lack of significance for ripple density implies that the short-time spectral changes are highly correlated across ripple density rate. Thus measuring the change over a narrow range of ripple densities can provide information related to the changes over a much wider range of densities, and a set of measurements covering the entire range of human spectral resolution is not needed for predicting changes in intelligibility or quality.

The information analysis indicates that the relative importance of low as compared to high modulation rates for intelligibility is similar for NH and HI listeners once the peripheral loss is taken into account. The relative importance of different auditory bands or spectral ripple terms is also similar for the two groups of listeners. Even though the HI listeners have poorer intelligibility, the relative importance of the different envelope modulation rates in providing usable speech information is similar to that of the NH listeners once the peripheral loss is taken into account by the auditory model.

Both the NH and HI groups of listeners gave similar quality ratings for both degraded speech and music signals (Arehart *et al.*, 2010, 2011), and again the relative importance of the low and high modulation rates and the dependence on auditory frequency is similar for the two groups. Thus, given an accurate model of the auditory periphery including the effects of hearing loss, the same basic models can be used for both NH and HI listeners to accurately predict intelligibility or quality (Kates and Arehart, 2014a,b) averaged over the subject groups. Accounting for individual variability, however, may require individual tuning of the peripheral model or developing perceptual models that extend higher up the auditory pathway (Kates *et al.*, 2013).

While the uncertainty coefficients indicate the relative information provided by the different modulation rates and auditory frequencies for intelligibility and quality, they are not in themselves the weights that should be applied in constructing an index. In building an index, one must consider not only the relative importance of each modulation rate and auditory band, but also the potential duplication of information between different modulation rates and auditory bands as well as the functional relationship between the measured signal quantities and the subject scores or ratings being modeled. The uncertainty coefficients thus provide a general guide as to which modulation rates and auditory frequency bands may be the most useful in constructing an index.

V. CONCLUSIONS

This paper used mutual information to quantify the relationship between changes in envelope modulation rate and perceptual responses to degraded speech and music signals. A model of the auditory periphery was used to generate the envelope signals, and the calculations were based on the normalized cross-covariance of the degraded signal envelope with that of a reference signal in each modulation frequency band. The envelope comparisons thus measured envelope modulation fidelity rather than intensity. Responses from previous intelligibility, speech quality, and music quality experiments were evaluated. The uncertainty coefficients were computed over all of the processed stimuli present in each dataset, which provided an average over a wide range of signal conditions.

The results showed that the uncertainty coefficients for envelope modulation rate were a significant factor for modulation evaluated within auditory analysis bands and for a spectro-temporal modulation analysis. The results also showed that auditory analysis frequency was a significant factor in the statistical analysis, but that differences in the information conveyed by the different cepstral correlation basis functions were in general not significant. Changes at low envelope modulation rates were found to convey the most information for intelligibility, while changes at high modulation rates were most important for speech and music quality. Low-to-mid auditory frequencies were most important for intelligibility, while mid frequencies were most important for speech quality and high frequencies were most important for music quality.

Accurate intelligibility and quality indices have been built based on both of the approaches used in this study for envelope modulation analysis. In general, these indices average over modulation rate and/or auditory frequency. For procedures that use envelope modulation analyzed within auditory frequency bands, the results suggest that improved accuracy may be possible by using rate-dependent and frequency-dependent weights in the calculations. For procedures based on cepstral correlation, the results suggest a benefit for using rate-dependent weights. For both approaches, the weights should be adjusted depending on whether the index is intended for intelligibility, speech quality, or music quality.

The results showed similar patterns for NH and HI listener groups when assessing the information contained at different modulation rates. Because the envelope analysis used the output of an auditory model, it included many of the signal modifications, such as auditory threshold shift and broader auditory filters, which are introduced by the hearing loss. The measurement of the envelope fidelity at the output of the HI auditory model appears to provide information for the HI listeners comparable to that provided by the normal-hearing auditory model for the NH listeners when the results are averaged over the subject groups. The information calculations indicate that the same relative amount of information is being provided for both hearing groups, but it does not necessarily mean that the information is being used in the same way and it does not completely account for individual variability within the groups. The practical result, however, is that the same intelligibility or quality index may be used to predict performance for both normal-hearing and hearing-impaired subjects given an adequate model of the auditory periphery.

ACKNOWLEDGMENTS

The authors thank Alexandre Chabot-Leclerc and Torsten Dau for providing the modulation filter band weights used in their recent paper (Chabot-Leclerc *et al.*, 2014) and for providing comments on a draft version of this paper. J.M.K. was supported by a research grant from GN ReSound. K.H.A. was supported by a NIH Grant No. R01 DC012289 and by the grant from GN ReSound.

Aguilera Muñoz, C. M., Nelson, P. B., Rutledge, J. C., and Gago, A. (**1999**). "Frequency lowering processing for listeners with significant hearing loss," in *Proceedings of Electronics, Circuits, and Systems: ICECS1999*, Pafos, Cypress, September 5–8, 1999, Vol. 2, pp. 741–744.

Anderson, M. C. (**2010**). "The Role of Temporal Fine Structure in Sound Quality Perception," Ph.D. thesis, Department of Speech, Language, and Hearing Sciences, University of Colorado, Boulder, 2010.

Arehart, K. H., Kates, J. M., and Anderson, M. C. (**2010**). "Effects of noise, nonlinear processing, and linear filtering on perceived speech quality," Ear Hear. **31**, 420–436.

Arehart, K. H., Kates, J. M., and Anderson, M. C. (**2011**). "Effects of noise, nonlinear processing, and linear filtering on perceived music quality," Int. J. Audiol. **50**, 177–190.

Arehart, K. H., Souza, P., Baca, R., and Kates, J. M. (**2013a**). "Working memory, age, and hearing loss: Susceptibility to hearing aid distortion," Ear Hear. **34**, 251–260.

Arehart, K. H., Souza, P. E., Lunner, T., Pedersen, M. S., and Kates, J. M. (**2013b**). "Relationship between distortion and working memory for digital noise-reduction processing in hearing aids," POMA **19**, 050084.

Başkent, D. (**2006**). "Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels," J. Acoust. Soc. Am. **120**, 2908–2925.

Benesty, J., Chen, J., and Huang, Y. (**2008**). "On the importance of the Pearson correlation coefficient in noise reduction," IEEE Trans. Audio Speech Lang. Process. **16**, 757–765.

Byrne, D., and Dillon, H. (**1986**). "The National Acoustics Laboratories' (NAL) new procedure for selecting gain and frequency response of a hearing aid," Ear Hear. **7**, 257–265.

Chabot-Leclerc, A., Jørgensen, S., and Dau, T. (**2014**). "The role of auditory spectro-temporal modulation filtering and the decision metric for speech intelligibility prediction," J. Acoust. Soc. Am. **135**, 3502–3512.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectrotemporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Christiansen, C., Pedersen, M. S., and Dau, T. (**2010**). "Prediction of speech intelligibility based on an auditory preprocessing model," Speech Commun. **52**, 678–692.

Croghan, N. B. H., Arehart, K. H., and Kates, J. M. (**2014**). "Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics," Ear Hear. **35**, e170–e184.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of reducing slow temporal modulations on speech perception," J. Acoust. Soc. Am. **95**, 2670–2680.

Dubbelboer, F., and Houtgast, T. (**2008**). "The concept of the signal-to-noise ratio in the modulation domain and speech intelligibility," J. Acoust. Soc. Am. **124**, 3937–3946.

Elhilali, M., Chi, T., and Shamma, S. (**2003**). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," Speech Commun. **41**, 331–348.

Falk, T. H., Zheng, C., and Chan, W.-Y. (**2010**). "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," IEEE Trans. Audio Speech Lang. Process. **18**, 1766–1774.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Goldsworthy, R. L., and Greenberg, J. E. (**2004**). "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," J. Acoust. Soc. Am. **116**, 3679–3689.

Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Houtgast, T., and Steeneken, H. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

Huber, R., and Kollmeier, B. (**2006**). "PEMO-Q: A new method for objective audio quality assessment using a model of auditory perception," IEEE Trans. Audio Speech Lang. Process. **14**, 1902–1911.

International Telecommunication Union (**2003**). ITU-R BS.1284-1, "General methods for the subjective assessment of sound quality."

Jørgensen, S., and Dau, T. (**2011**). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," J. Acoust. Soc. Am. **130**, 1475–1487.

Kates, J. M. (**2013**). "An auditory model for intelligibility and quality predictions," POMA **19**, 050184.

Kates, J. M., and Arehart, K. H. (**2005**). "Multichannel dynamic-range compression using digital frequency warping," EURASIP J. Appl. Sign. Process. **2005**, 3003–3014.

Kates, J. M., and Arehart, K. H. (**2010**). "The hearing aid speech quality index (HASQI)," J. Audio Eng. Soc. **58**, 363–381.

Kates, J. M., and Arehart, K. H. (**2014a**). "The hearing aid speech quality index (HASQI), version 2," J. Audio Eng. Soc. **62**, 99–117.

Kates, J. M., and Arehart, K. H. (**2014b**). "The hearing-aid speech perception index (HASPI)," Speech Commun. **65**, 75–93.

Kates, J. M., Arehart, K. H., and Souza, P. E. (**2013**). "Integrating cognitive and peripheral factors in predicting hearing-aid processing effectiveness," J. Acoust. Soc. Am. **134**, 4458–4469.

Kjems, U., Boldt, J. B., Pedersen, M. S., and Wang, D. (**2009**). "Role of mask pattern in intelligibility of ideal binary-masked noisy speech," J. Acoust. Soc. Am. **126**, 1415–1426.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acustica **83**, 659–669.

Li, N., and Loizou, P. C. (**2008**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 1673–1682.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

McAulay, R. J., and Quatieri, T. F. (**1986**). "Speech analysis/synthesis based on a sinusoidal representation," IEEE Trans. Acoust. Speech Sign. Process. **34**, 744–754.

Moddemeijer, R. (**1999**). "A statistic to estimate the variance of the histogram-based mutual information estimator based on dependent pairs of observations," Sign. Process. **75**, 51–63.

Nilsson, M., Ghent, R. M., and Bray, V. (**2005**). "Development of a test environment to evaluate performance of modern hearing aid features," J. Am. Acad. Audiol. **16**, 27–41.

Nilsson, M., Soli, S. D., and Sullivan, J. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Payton, K., and Shrestha, M. (**2013**). "Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data," J. Acoust. Soc. Am. **134**, 3818–3827.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, P. (**2007**). *Numerical Recipes 3rd Edition: The Art of Scientific Computing* (Cambridge University Press, London), p. 761.

Quatieri, T. F., and McAulay, R. J. (**1986**). "Speech transformations based on a sinusoidal representation," IEEE Trans. Acoust. Speech Sign. Process. **ASSP-34**, 1449–1464.

Rosenthal, S. (**1969**). "IEEE: Recommended practices for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246.

Schwartz, D. A., and Purvis, D. (**2004**). "Pitch is determined by naturally occurring periodic sounds," Hear. Res. **194**, 31–46.

Shannon, R. V., Zeng, F-Z., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Souza, P. E., Arehart, K. H., Kates, J. M., Croghan, N. B. H., and Gehani, N. (**2013**). "Exploring the limits of frequency lowering," J. Speech Lang. Hear. Res. **56**, 1349–1363.

Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.

Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J. (**2011**). "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," IEEE Trans. Audio Speech Lang. Process. **19**, 2125–2136.

Taghia, J., and Martin, R. (**2014**). "Objective intelligibility measures based on mutual information for speech subjected to speech enhancement processing," IEEE Trans. Audio Speech Lang. Process. **22**, 6–16.

Terhardt, E. (**1974**). "On the perception of periodic sound fluctuations (roughness)," Acoustica **30**, 201–213.

Tsoukalis, D. E., Mourjopoulos, J. N., and Kokkinakis, G. (**1997**). "Speech enhancement based on audible noise suppression," IEEE Trans. Speech Audio Process. **5**, 497–514.

Tufts, J. B., and Molis, M. R. (**2007**). "Perception of roughness by listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **121**, EL161–EL167.

Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.

van Buuren, R. A., Festen, J. M., and Houtgast, T. (**1999**). "Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality," J. Acoust. Soc. Am. **105**, 2903–2913.

van Veen, T. M., and Houtgast, T. (**1985**). "Spectral sharpness and vowel dissimilarity," J. Acoust. Soc. Am. **77**, 628–634.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based on modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Xu, L., and Pfingst, B. E. (**2008**). "Spectral and temporal cues for speech recognition: Implications for auditory prostheses," Hear. Res. **242**, 132–140.

Zahorian, S. A., and Rothenberg, M. (**1981**). "Principal-components analysis for low-redundancy encoding of speech spectra," J. Acoust. Soc. Am. **69**, 832–845.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models*, 2nd ed. (Springer-Verlag, New York), pp. 257–264.